Métodos de aprendizaje automático para predecir el comportamiento epidemiológico de enfermedades arbovirales: revisión estructurada de literatura

Machine learning methods to predict epidemiological behavior of arbovirals diseases: structured literature review

Sonia Isabel Polo-Triana[1]; Yuly Andrea Ramírez-Sierra[1]; Javier Eduardo Arias-Osorio[1]; Ruth Aralí Martínez-Vega[2]; Henry Lamos-Díaz[1]

*sonia2198627@correo.uis.edu.co

Forma de citar: Polo Triana SI, Ramírez Sierra YA, Arias Osorio JE, Martínez Vega A, Lamos Díaz H. Métodos de aprendizaje automático para predecir el comportamiento epidemiológico de enfermedades arbovirales: Revisión estructurada de literatura. Salud UIS. 2023; 55: e23017. doi: https://doi.org/10.18273/saluduis.55.e:23017

Resumen

Introducción: los métodos de aprendizaje automático permiten manejar datos estructurados y no estructurados para construir modelos predictivos y apoyar la toma de decisiones. **Objetivo:** identificar los métodos de aprendizaje automático aplicados para predecir el comportamiento epidemiológico de enfermedades arbovirales utilizando datos de vigilancia epidemiológica**. Metodología:** se realizó búsqueda en EMBASE y PubMed, análisis bibliométrico y síntesis de la información. **Resultados:** se seleccionaron 41 documentos, todos publicados en la última década. La palabra clave más frecuente fue dengue. La mayoría de los autores (88,3 %) participó en un artículo de investigación. Se encontraron 16 métodos de aprendizaje automático, el más frecuente fue Red Neuronal Artificial, seguido de Máquinas de Vectores de Soporte. **Conclusiones:** en la última década se incrementó la publicación de trabajos que pretenden predecir el comportamiento epidemiológico de arbovirosis por medio de diversos métodos de aprendizaje automático que incorporan series de tiempo de los casos, variables climatológicas, y otras fuentes de información de datos abiertos.

Palabras clave: Revisión; Infecciones por arbovirus; Vigilancia en salud pública; Predicción; Aprendizaje automático; Bibliometría.

[1] Universidad Industrial de Santander. Bucaramanga, Colombia.
[2] Universidad de Santander. Bucaramanga, Colombia.

Recibido: 06/09/2021 **Aprobado:** 05/10/2022

Abstract

Introduction: Machine learning methods allow to manipulate structured and unstructured data to build predictive models and support decision-making. **Objective:** To identify machine learning methods applied to predict the epidemiological behavior of vector-borne diseases using epidemiological surveillance data. **Methodology:** A literature search in EMBASE and PubMed, bibliometric analysis, and information synthesis were performed. **Results:** A total of 41 papers were selected, all of them were published in the last decade. The most frequent keyword was dengue. Most authors (88.3 %) participated in a research article. Sixteen machine learning methods were found, the most frequent being Artificial Neural Network, followed by Support Vector Machines. **Conclusions:** In the last decade there has been an increase in the number of articles that aim to predict the epidemiological behavior of vector-borne diseases using by means of various machine learning methods that incorporate time series of cases, climatological variables, and other sources of open data information.

Keywords: Review; Arboviral infections; Public health surveillance; Forecasting; Machine learning; Bibliometrics.

Introducción

Según la Organización Mundial de la Salud, las enfermedades transmitidas por vectores representan el 17 % de todas las enfermedades infecciosas en el mundo, generan más de 700 000 muertes cada año, tienen la mayor carga en áreas tropicales y subtropicales, y afectan a las poblaciones más pobres[1]. Entre estas se encuentran las enfermedades arbovirales, que continúan siendo un problema de salud pública mundial[2] y son ocasionadas por virus que se transmiten entre hospederos vertebrados por medio de artrópodos, como mosquitos, garrapatas, entre otros. El dengue continúa siendo la enfermedad arboviral de mayor frecuencia en el mundo. Sin embargo, la reemergencia y emergencia de otras enfermedades como fiebre amarilla, chikungunya y zika, todas estas transmitidas por mosquitos *Aedes* spp, acentúa que las enfermedades arbovirales siguen siendo un problema de salud en el siglo XXI y que se requiere reevaluar las prioridades de investigación y las intervenciones de salud pública contra estas enfermedades[3].

En Colombia, desde su reemergencia en la década de los 70, el dengue ha tenido un comportamiento endemo-epidémico, con epidemias en la última década en 2010, 2013, 2016 y 2019 y alrededor del 80 % de los municipios del país reportan casos de dengue al sistema de vigilancia epidemiológica. Además, el país ha transitado por la emergencia de chikungunya en 2014 y de zika en 2015[4]; y se ha evaluado que 87,6 % de las localidades colombianas tienen presencia del vector *Aedes aegypti*[5]. También, se ha estimado que estas tres enfermedades causaron 491 629,2 años de vida ajustados por discapacidad (AVAD) entre 2013 y 2016, siendo superior a la carga ocasionada en algunos años de este periodo por la tuberculosis y VIH/SIDA[6].

Estas enfermedades pueden prevenirse a través de la combinación de diferentes estrategias de salud pública[3], entre estas se encuentran la vigilancia epidemiológica y la vigilancia vectorial, así como el desarrollo de modelos predictivos de incidencia que utilicen aproximaciones de diversas disciplinas como la epidemiología y el aprendizaje automático[2]. El valor de la vigilancia radica en la entrega oportuna de datos útiles, la flexibilidad de los sistemas al incremento en las necesidades de información, así como en el uso apropiado de las tecnologías para difundirla oportunamente. Por tanto, la vigilancia en salud pública puede beneficiarse de los avances en ciencias de la información, de las tecnologías, y del aumento de fuentes y bases de datos[7].

En este sentido, emerge un área interdisciplinar denominada epidemiología computacional, que utiliza modelos computacionales y big data para entender y controlar la difusión espacio-temporal de una enfermedad en la población[8]. De manera que se está optando por desarrollar tecnologías en bases de datos, así como aplicar técnicas de analítica como minería de datos y aprendizaje automático, que permiten manejar información estructurada y no estructurada, construir modelos predictivos con variedad de técnicas basados en datos históricos, y apoyar la toma de decisiones junto con herramientas y técnicas de visualización[9–11]. Por lo anterior, el objetivo de la presente revisión fue identificar los métodos de aprendizaje automático utilizados con el fin de predecir el comportamiento epidemiológico, entendido este como la frecuencia de ocurrencia de casos incidentes de enfermedades arbovirales en un periodo de tiempo específico, así como documentar las diferentes variables predictoras usadas.

Metodología

Pregunta de investigación: ¿Cuáles son los métodos de aprendizaje automático que se han utilizado para predecir el comportamiento epidemiológico de las enfermedades arbovirales?

Fuentes de información y estrategia de búsqueda: los documentos se obtuvieron por medio de EMBASE y PubMed haciendo uso de la siguiente ecuación: ((“predictive model*” OR predict* OR forecast) AND (“computational epidemiology” OR “descriptive analyt*” OR “predictive analyt*” OR “machine learning” OR “deep learning” OR “supervised classifi*” OR “supervised* algorithms” OR “supervised learning” OR “unsupervised classifi*”)) AND (“arboviral diseas*” OR arbovirus* OR “vector-borne diseas*” OR “transmitted by vector*” OR dengue OR zika or chikungunya).

Además de los artículos obtenidos a través de la ecuación, se incluyeron algunos por medio del principio de bola de nieve.

Criterios de elegibilidad: se incluyeron artículos que utilizaron el aprendizaje automático para la predicción del comportamiento epidemiológico de enfermedades arbovirales. Se excluyeron los artículos enfocados en la predicción de desenlaces individuales y de clases de virus, o que no aportaran información de interés a esta revisión. No se realizó filtro por idioma o por fecha.

Síntesis de la información: se adoptó el flujo de trabajo PRISMA[12]. La ecuación de búsqueda arrojó 413 artículos. Se agregaron 6 artículos utilizando el principio de bola de nieve. Se obtuvieron 346 artículos después de la eliminación de duplicados. Inicialmente, se realizó un filtro por lectura del título y resumen de cada artículo, conservando 96 artículos relevantes. Después, la lectura de texto completo permitió seleccionar 41 artículos que cumplieron con los criterios de elegibilidad (Figura 1). Además, se realizó un análisis bibliométrico utilizando la nube de palabras, la co-ocurrencia entre las palabras clave, las coautorías y las revistas en las cuales se publicaron los artículos seleccionados.

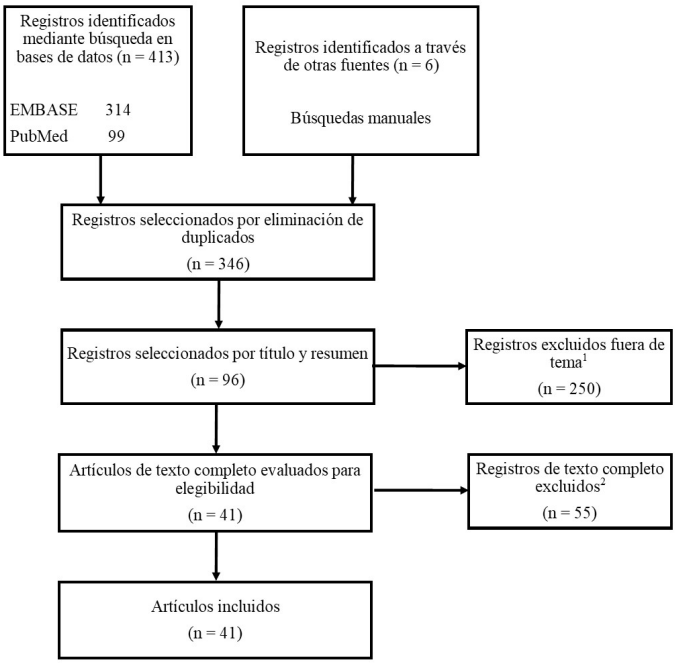

Figura 1. Flujograma de la búsqueda de la literatura realizada.

1. Exclusiones fuera de tema: Estudios del genoma o proteínas del virus, ontologías y taxonomías, estudios de inhibidores, estudios de predicción a nivel individual (diagnóstico de enfermedad o necesidad de hospitalización, predicción de la gravedad de la enfermedad), modelación del vector o predicción de la abundancia del vector, desarrollo de vacunas o de fármacos, redes de información (de salud), predicción del tipo de arbovirus o parásitos, diseño de sistemas de vacunación, enfermedades diferentes a las transmitidas por vectores.

2. Exclusiones de texto completo: No contenían la aplicación de métodos de aprendizaje automático. Artículos o estudios en los que el modelo se entrenó para mejorar los resultados de los sistemas de gestión de la atención médica. Se excluyeron los artículos cuyo objeto de estudio fuese diferente a la predicción del número de casos, nivel de incidencia o la probabilidad de ocurrencia de la enfermedad arboviral. Se excluyeron los artículos sin acceso a texto completo con las licencias institucionales, o estudios que fuesen revisiones de literatura.

Resultados

Análisis bibliométrico

Los 41 artículos seleccionados a partir de la búsqueda se publicaron en la última década, se observó un incremento en la publicación a partir de 2016 (Figura 2). El 88,3 % de los autores participó en un artículo, mientras que el 8,9 % participó en dos artículos y el 2,8 % restante en tres artículos. En cuanto a la revista de publicación, la más frecuente fue PLoS Neglected Tropical Diseases (24,4 %), seguida por PLoS One (3,3 %) y BMC Infectious Diseases (3,3 %). Adicionalmente, en Scientific Reports, Acta Tropica, BMC Medicine y Epidemiology and Infection se publicaron dos artículos en cada una.

En relación con los intervalos de tiempo considerados en los artículos, 1 estudio usó casos informados diariamente, 33 estudios casos semanales, 1 estudio usó casos mensuales y 6 estudios no especificaron el intervalo de tiempo. Además, la serie de tiempo más larga fue de 27 años, mientras que la más corta fue de 1 año, con un promedio de 8 años y una mediana de 5 años. En 24 estudios se utilizaron series de tiempo de menos de 10 años, mientras que en 11 estudios la serie de tiempo fue de 10 años o más.

Por otro lado, la palabra clave más frecuente fue dengue, seguida de pronóstico (forecasting), epidemiología (epidemiology) y zika. Después, figuraron los factores asociados al comportamiento epidemiológico de las arbovirosis como humedad relativa, temperatura y el vector (*Aedes aegypti, Aedes albopictus*). Llama la atención que los métodos analíticos aparecen en el grupo de palabras menos frecuentes (Figura 3). Al analizar la co-ocurrencia entre palabras clave se evidenció como nodo principal a dengue relacionado con diversos métodos analíticos como redes neuronales artificiales, LASSO, Fuzzy clustering, redes bayesianas, entre otros. Además, el dengue se relacionó con los factores asociados al comportamiento epidemiológico; en contraste, zika se relacionó con machine learning, mapas de riesgo y epidemiología (Figura 4). Los autores con más artículos fueron Lee Ching Ng y Jayanthi Rajarethinam, que participaron en cuatro, y Ai-ping, Deng, Tie Song, Meng Zhang, Alex Cook, Shaohong, Liang, Xu Liu y Grace Yap que participaron, cada uno, en tres artículos. Al analizar la coautoría se observó mayor interrelación entre los autores del periodo 2018 - 2019, con una posterior concentración en 2020 (Figura 5). Es importante aclarar que algunos de los autores con mayor cantidad de artículos, no figuran en la gráfica de coautorías por que publicaron con los mismos coautores, y esta gráfica refleja la cantidad de coautores diferentes con los que ha publicado el investigador, mas no la cantidad de artículos.

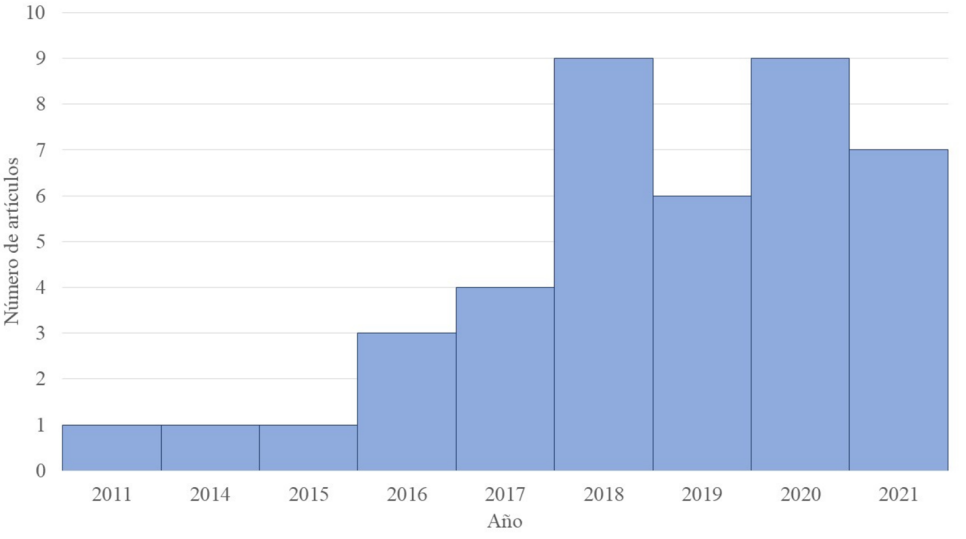

Figura 2. Publicación de los artículos incluidos la revisión de la literatura por año.

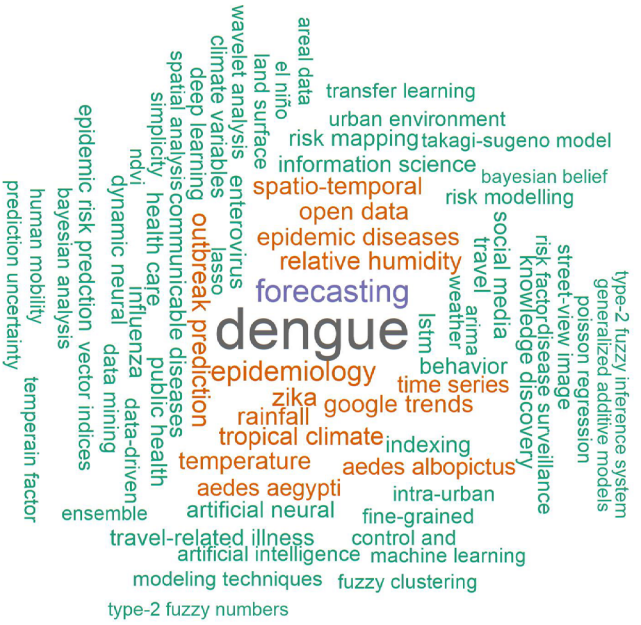

Figura 3. Nube de palabras clave de los artículos incluidos en la revisión.

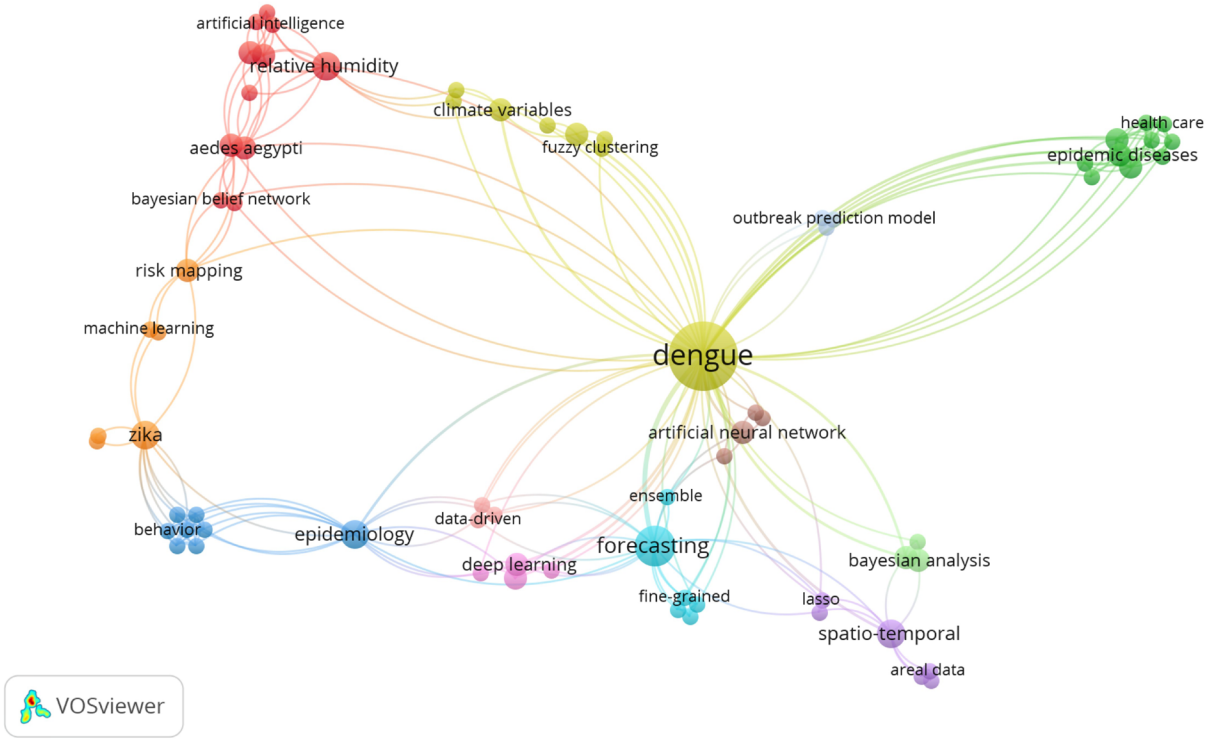

Figura 4. Co-ocurrencia de palabras clave de los artículos incluidos en la revisión.

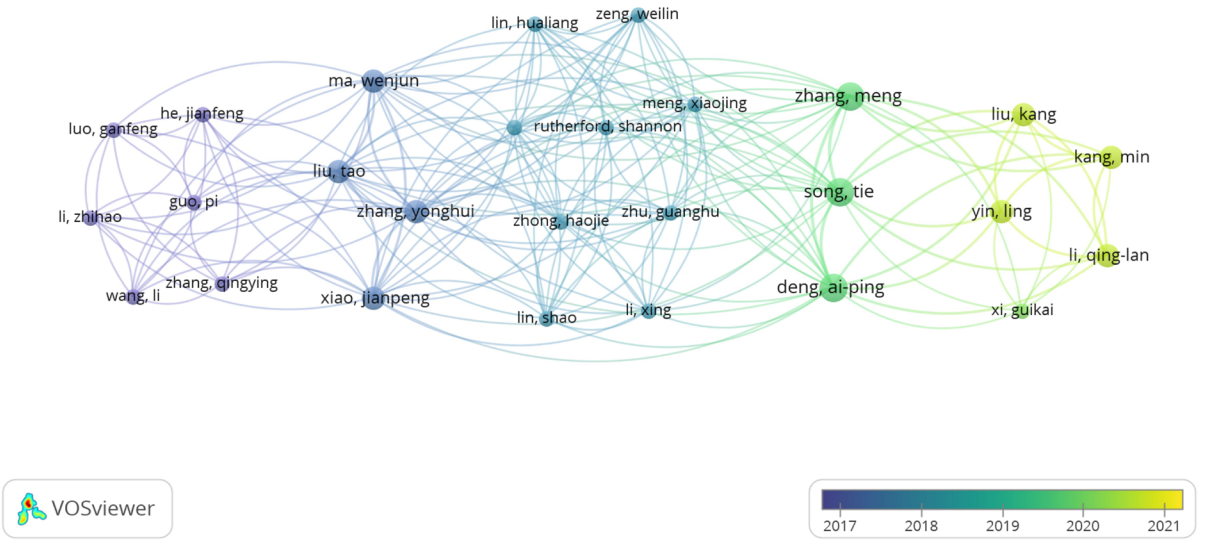

Figura 5. Co-autoría de los artículos incluidos en la revisión.

Herramientas para la vigilancia de enfermedades arbovirales como fuentes de datos para el uso de aprendizaje automático

La vigilancia epidemiológica consiste en "la recopilación, el análisis y la interpretación, sistemática y continua de datos de salud esenciales para la planificación, implementación y evaluación de la práctica de salud pública, estrechamente integrada con la diseminación oportuna de estos datos a quienes necesitan conocerlos"[13]. Dentro de las herramientas para la vigilancia, además de la información sobre los casos de la enfermedad, los datos de la vigilancia entomológica son muy importantes porque pueden facilitar la comprensión de la ecología de los vectores en un área determinada y permiten determinar el efecto de las estrategias de intervención anti-vectorial[14,15].

Otra herramienta para la vigilancia son los sistemas de información geográfica, que ayudan a identificar casos y exposiciones, tendencias espaciales, así como a correlacionar diferentes conjuntos de datos y probar hipótesis estadísticas. Actualmente, el desarrollo de sistemas cada vez más sofisticados ayuda a los profesionales de la salud pública a monitorear y responder a los desafíos de salud[16]. Por ejemplo, las herramientas para georreferenciar permiten analizar las relaciones espaciales entre áreas con altos niveles de infestación de mosquitos y los sitios óptimos para el desarrollo larvario, y ahorran tiempo en la identificación de áreas problemáticas para que los trabajadores de salud pública realicen intervenciones de control[17]. La epidemiología molecular es otra herramienta que se ocupa de comprender la distribución y las relaciones de variantes genéticas, serotipos u otras agrupaciones moleculares de los patógenos, visualizándolas con árboles filogenéticos o dendrogramas[18]. Los análisis espacio-temporales de los genomas se centran en el mapeo y la predicción del intercambio de linajes de virus entre ubicaciones, para reconstruir las vías de introducción y propagación[19]. Además, estas herramientas ayudan a caracterizar el número de introducciones de un patógeno que conducen a la transmisión en una nueva ubicación, cuantificar el riesgo de transmisión entre especies y a inferir factores ecológicos de transmisión[20].

Por otra parte, en la literatura sobre salud pública se evidencia el incremento de estudios que utilizan el análisis de redes sociales como otra herramienta para la vigilancia; los propósitos se centran en la identificación de las características comunes de los infectados, la comunicación visual o el mapeo de los casos para una mejor comprensión de los brotes, y la identificación de posibles vías de transmisión y características geográficas[21,22,23]. Los estudios han encontrado que los usuarios discuten públicamente los síntomas y comportamientos utilizados para prevenir dolencias[24,25]. Sin embargo, se conoce que el uso de las redes sociales tiene un sesgo demográfico. Por

ejemplo, en Brasil utilizaron tweets de eventos de dengue geoetiquetados para explorar las tendencias de notificación espacio-temporales, y los relacionaron con los casos confirmados. Sus resultados mostraron que el volumen de tweets fue significativamente menor que el volumen de casos de dengue, y lo asociaron a la desigualdad y bajo ingreso de estas poblaciones, donde existe poca cultura para el uso de ciertas tecnologías en el área de la salud[26].

La teledetección es otra herramienta para la vigilancia, puesto que los satélites diariamente recopilan grandes cantidades de datos y con diferentes tipos de medición que se pueden utilizar para desarrollar modelos predictivos para la propagación de enfermedades transmitidas por vectores[27]. Por ejemplo, se han hallado correlaciones entre el riesgo de transmisión del virus del Nilo Occidental y los cambios en la floración primaveral, la temperatura y la humedad, observados por satélites en EE. UU[28]. Además, se pueden obtener diversos tipos de indicadores e informar sobre los hábitats y riesgos de enfermedades transmitidas por mosquitos, garrapatas, moscas negras y flebótomos[29]. También se ha reportado que, en los países donde la vigilancia es escasa o está fracasando, el uso de datos de teledetección ha contribuido a la predicción de enfermedades emergentes transmitidas por vectores, permitiendo fomentar y focalizar medidas preventivas, como el control de vectores en las zonas afectadas[27,30].

Todas estas herramientas generan información que, junto con el recuento de casos de la vigilancia, ha sido utilizada para tratar de explicar o predecir el comportamiento epidemiológico de las enfermedades arbovirales e intentar disminuir su impacto con el uso de diferentes metodologías de análisis.

Métodos de aprendizaje automático usados para predecir el comportamiento epidemiológico de enfermedades arbovirales

En los 41 artículos incluidos en esta revisión se encontraron 16 métodos de aprendizaje automático diferentes. La red neuronal artificial fue el más utilizado (n = 12), seguido de las máquinas de vectores de soporte (n = 9), las redes bayesianas (n = 8), los bosques aleatorios (n = 8), la regresión LASSO (n = 7), la media móvil integrada autorregresiva o media móvil integrada autorregresiva estacional (ARIMA/SARIMA, por sus siglas en inglés) (n = 7) y el modelo aditivo generalizado (GAM, por sus siglas en inglés) (n = 6) (Tabla 1).

Tabla 1. Métodos de aprendizaje automático que se han utilizado para predecir el comportamiento epidemiológico de enfermedades arbovirales

Método	Fuente	Cantidad de estudios
Red Neuronal Artificial	Salim et al. 2021[43]; Liu et al. 2021[64]; Mussumeci & Coelho, 2020[44]; Bomfim et al. 2020[44]; Zhao et al. 2020[39]; Polwiang, 2020[57]; Ho et al., 2020[67]; Xu et al. 2020[45]; Liu et al. 2020[40]; Akhtar et al. 2019[47]; Jiang et al. 2018[42]; Baquero et al. 2018[32]	12
SVM	Nejad & Varathan, 2021[60]; McGough et al.2021[61]; Liu et al. 2021[64]; Xu et al. 2020[45]; Liu et al. 2020[40]; Stolerman et al. 2019[52]; Guo et al. 2017[31]; Nsoesie et al. 2016[26]; Althouse et al. 2011[22]	9
Redes Bayesianas	Salim et al. 2021[43]; Nejad & Varathan, 2021[60]; Akter et al. 2021[35]; Aswi, et al. 2019[34]; Baquero et al. 2018[32]; Martínez-Bello et al. 2017[37]; Ho et al. 2017[36]; Nsoesie et al. 2016[26]	8
Random Forest	Mussumeci & Coelho, 2020[44]; Benedum et al. 2020[62]; Zhao et al. 2020[39]; Carvajal et al. 2018[63]; Jiang et al. 2018[42]; Ong et al. 2018[56]; Xiao et al. 2018[68]; Liu et al. 2016[65]	8
Regresión LASSO	Mussumeci & Coelho, 2020[44]; Liu et al. 2020[40]; Chen et al. 2018[53]; Wu et al. 2018[30]; Guo et al. 2017[31]; McGough et al. 2017[25]; Shi et al. 2016[54]	7
ARIMA // SARIMA	Benedum et al. 2020[62]; Bomfim et al. 2020[46]; Zhao et al. 2020[39]; Polwiang, 2020[57]; Carvajal et al. 2018[63]; Baquero et al. 2018[32]; Teng et al. 2017[23]	7
GAM	Xu et al. 2020[45]; Jain et al. 2019[33]; Carvajal et al. 2018[63]; Xiao et al. 2018[68]; Baquero et al. 2018[32]; Guo et al. 2017[31]	6
Árboles de decisión	Salim et al. 2021[43]; Wu & Kao, 2021[51]; Nejad & Varathan, 2021[60]; Ho et al. 2020[67]; Campbell et al. 2015[55]	5
GBM	Xu et al. 2020[45]; Tuladhar et al. 2019[58]; Carvajal et al. 2018[63]; Jiang et al. 2018[42]	4
Regresión Logística	Ho et al. 2020[67]; Daughton & Paul, 2019[24]; Benedum et al. 2018[15]; Althouse et al. 2011[22]	4

Método	Fuente	Cantidad de estudios
NBM	Liu et al. 2021[64]; Guo et al. 2017[31]; Althouse et al. 2011[22]	3
Modelos Difusos	Adak & Jana, 2021[50]; Torres et al. 2014[49]	2
GLM	Yuan et al. 2020[59]; Ho et al., 2020[67]	2
SEIR Modelo Mecanicista	Bomfim et al. 2020[46]	1
Regresión de Poisson	Polwiang, 2020[57]	1
Máxima Entropía	Nsoesie et al. 2016[26]	1

SVM = Support Vector Machines (Máquina de vectores de soporte)
GLM = Generalized Linear Models (Modelos lineales generalizados)
GBM = Gradient Boosting Machine (Máquina de aumento de gradiente)
GAM = Generalized Additive Model (Modelo aditivo generalizado)
NBM = Negative Binomial Regression Model (Modelo de regresión binomial negativa)

Algunos de estos estudios utilizaron varios modelos para predecir el comportamiento epidemiológico y los compararon para determinar cuál de ellos tenía el mejor desempeño. Por ejemplo, Guo et al. evaluaron seis algoritmos para predecir la incidencia de dengue en China y encontraron que el modelo GAM obtuvo los valores del promedio del error cuadrático medio (RMSE) más altos, mientras que el modelo de regresión de vectores de soporte (SVR) los tuvo más bajos en comparación con los modelos de regresión binomial negativa (NBM), la máquina de aumento de gradiente (GBM) y la regresión LASSO (least absolute shrinkage and selection operator), por tanto, el SVR fue considerado el modelo de mejor desempeño[31]. Por otro lado, en São Paulo, Baquero et al. realizaron un estudio comparativo de las predicciones del dengue utilizando diferentes métodos. Ellos encontraron que el GAM fue el más preciso, pues predijo correctamente grandes epidemias, incluidos los picos de 2014 y 2015, con un mejor desempeño que los modelos de red neuronal artificial (ANN) y SARIMA[32].

Una de las aplicaciones más recientes fue realizada por Jain et al, quienes por medio de datos meteorológicos, clínicos, socioeconómicos y de vigilancia, desarrollaron modelos GAM para ajustar las relaciones entre los predictores (con un rezago de un mes) y el desenlace de brote de dengue hemorrágico. La evaluación de la capacidad de discriminación del modelo final contra el umbral constante específico de Bangkok y el umbral móvil de la epidemia de la OMS logró una especificidad del 92,6 %, sensibilidad del 87 %, un valor predictivo positivo del 95,7 % y un valor predictivo negativo del 78,8 %[33].

Adicional a los modelos predictivos, se encuentran los modelos espaciales bayesianos, los cuales son usados para el análisis espacial de enfermedades porque pueden reducir la varianza estimada de la variable respuesta, particularmente para regiones con poblaciones pequeñas. También incorporan una gama amplia de componentes de varianza en diferentes niveles en el modelo y es más fácil obtener una evaluación más completa de la incertidumbre de la predicción basada en la máxima verosimilitud[34]. Por ejemplo, Akter et al. utilizaron modelos de regresión de Poisson multivariados con enfoque bayesiano y estructura previa condicional autorregresiva para analizar la variación espacial de las notificaciones de dengue en relación con la variabilidad climática y los factores socioecológicos en la zona climática tropical de Queensland, Australia[35]. En otro estudio, Ho et al. presentaron una red de creencias bayesianas (BBN) para evaluar el riesgo potencial de aparición y distribución del virus dengue en Australia Occidental, mientras que Martínez-Bello et al. aplicaron modelos espacio-temporales bayesianos jerárquicos para evaluar el riesgo relativo del dengue en Colombia[36,37].

Es importante resaltar que el aprendizaje automático es una herramienta que combina la estadística con la informática mediante un uso eficiente de los conjuntos de datos masivos y de modelos no lineales con alta dimensionalidad[38,39,40]. Además, este se diferencia del modelo estadístico tradicional (por ejemplo, modelos de regresión) en que hay menos supuestos sobre la distribución subyacente de los datos y las relaciones entre las variables[41,42]. Por ejemplo, Salim et al. utilizaron varios modelos de aprendizaje automático para predecir los brotes de dengue, a saber: árboles de decisión (CART), ANN, SVM (Lineal, Polynomial, RBF) y Red Naïve Bayes (TAN). La variable de brote de dengue (1 = brote y 0 = sin brote) se creó con base en el número de casos de dengue notificados y se utilizaron

1300 registros. Estos modelos fueron similares en el sentido de que lograron obtener la probabilidad de la variable objetivo binaria e identificaron los predictores importantes[43].

En Brasil, Mussumeci & Coelho compararon un modelo de red neuronal de memoria a largo y corto plazo (LSTM), un modelo de regresión de bosque aleatorio (RF) y regresión LASSO para predecir la incidencia de dengue en 790 ciudades brasileñas. Además, considerando que es muy difícil pronosticar con precisión la incidencia semanal en una ciudad utilizando solo sus propios datos históricos y que la transmisión de dengue tiene un componente espacial relacionado con la movilidad de las personas entre ciudades, utilizaron series de los predictores agrupando ciudades vecinas. Para esto definieron el grupo de ciudades dentro de un mismo estado con base en las distancias de correlación y para cada ciudad se ajustaron y probaron los modelos. Los datos históricos correspondieron a series de tiempo semanal (442 semanas en total) de incidencia de dengue, temperatura, humedad relativa, presión atmosférica y tuits sobre el dengue de cada grupo de ciudades. Los modelos LSTM y RF tuvieron los errores de predicción más bajos, sin embargo, LSTM pudo aproximar mejor la distribución empírica de la incidencia semanal del dengue y fue el menos sesgado de los tres modelos que sobrestimaron ligeramente las incidencias muy bajas de periodos interepidémicos mientras que subestimaron los picos epidémicos[44].

Xu et al. propusieron un modelo basado en LSTM para predecir de manera eficiente los casos de dengue en 20 ciudades de China continental. Sus resultados mostraron que el LSTM redujo el promedio del RMSE de las predicciones de un 24,91 % a un 12,99 % y el RMSE promedio de las predicciones en el período del brote de un 26,82 % a 15,09 %, en comparación con GAM, SVR y GBM[45]. De manera similar, en un estudio en Fortaleza, Brasil, el LSTM pronosticó con precisión los casos de dengue, y la inclusión de datos de movilidad humana mejoró sustancialmente su desempeño[46].

Adicionalmente, Akhtar et al. aplicaron un modelo de red neuronal dinámica basada en modelos autorregresivos no lineales con entradas exógenas conocidas como redes neuronales NARX, haciendo uso de datos epidemiológicos, volúmenes de viajes aéreos de pasajeros, idoneidad del hábitat del vector *Ae. Aegypti,* y datos socioeconómicos y de población. Su modelo se basó en un clasificador de riesgo binario, es decir, clasificar una región k como de alto o bajo riesgo en el tiempo t + N. El modelo se entrenó y probó utilizando los datos disponibles hasta la semana (X – N). Por ejemplo, la predicción con 12 semanas de anticipación para la semana 40 se realizó usando los datos disponibles hasta la semana 28. La precisión del modelo estuvo por encima del 80 %[47].

Además, el sistema de inferencia difusa (FIS), una de las técnicas de inteligencia artificial más populares y aceptadas, se ha convertido recientemente en una potente herramienta para desarrollar y analizar muchos sistemas del mundo real. FIS consta de funciones de pertenencia, operador de lógica difusa y algunas reglas definidas por el usuario que están en formato "si-entonces"[48]. Al respecto, Torres et al. emplearon un enfoque metodológico que combinó análisis multirresolución y FIS para evaluar su capacidad predictiva respecto al número de casos de dengue y dengue severo en un horizonte de tres años en Colombia. Acorde con sus resultados, la técnica logró un rendimiento significativamente superior al obtenido con las técnicas tradicionales de modelado difuso utilizadas hasta el momento. De esta forma, la similitud entre los datos originales y la señal aproximada aumentó del 21,13 % al 90,06 % y del 18,90 % al 76,83 % en el caso del dengue y el dengue severo, respectivamente[49].

En una aplicación más reciente, Adak y Jana desarrollaron un modelo matemático utilizando un FIS tipo 2 para predecir las condiciones adecuadas para un brote de dengue, en el cual consideraron como variables de entrada los datos climatológicos (Temperatura, Precipitaciones y Humedad relativa) y como variable de salida la "Probabilidad del dengue". Para cada una de las variables se definieron rangos, mediante los cuales se asignaron las funciones de pertenencia que se tomaron como 'baja', 'media' y 'alta', consideraron todas las combinaciones posibles de las variables de entrada para asegurar la estabilidad del modelo. Los autores desarrollaron tres sistemas, el primer (R^2 = 0,8169) y segundo sistema (R^2= 0,6832) funcionaron satisfactoriamente para predecir la probabilidad de ocurrencia del dengue en contraste con el tercer sistema (R^2 = 0,6271), donde los valores predichos sobrestimaron los observados[50].

Finalmente, la extracción de conocimientos de múltiples fuentes de datos abiertos en el campo de las enfermedades epidémicas se ha vuelto cada vez más relevante para la gestión de la salud pública, particularmente en relación con enfermedades causadas por algunos virus como Dengue, por lo que Wu y Kao en su estudio aplicaron un enfoque de minería de datos en un modelo de predicción en el cual consideraron datos

de Google Trends. Sin embargo, al evaluar su efecto y desempeño en el modelo predictivo, no encontraron que esta información mejorara la predicción del dengue[51].

Discusión

Algunas enfermedades arbovirales tienen un comportamiento epidemiológico estacional, es decir presentan un patrón regular de aumento de casos en ciertos periodos del año. Por consiguiente, el manejo eficaz de enfermedades estacionales depende del despliegue oportuno de medidas de control antes de la temporada de alta transmisión. Dado que la temporada epidémica varía de un año a otro, la disponibilidad de pronósticos precisos de incidencia puede ser decisiva para lograr el control de tales enfermedades o la preparación de los sistemas de salud para atender los casos que se presenten de tal forma que se disminuya la morbimortalidad[44,52]. Se han realizado múltiples investigaciones en diferentes partes del mundo para estudiar variables asociadas al comportamiento epidemiológico de las enfermedades arbovirales y se han tenido en cuenta diferentes fuentes de información y factores. Dentro de los factores estudiados se encuentran las condiciones socioeconómicas, el tamaño de la población, los comportamientos relacionados con el almacenamiento de agua, la urbanización, los índices vectoriales, la movilidad humana[53,54,55], datos entomológicos[56] y variables climatológicas como la temperatura, las precipitaciones, la humedad, y la velocidad del viento, entre otras[46,50,54,57–63].

Adicionalmente, la posibilidad de utilizar información de múltiples fuentes de datos abiertos con el objetivo de predecir enfermedades epidémicas[51,64,65], cuyas soluciones propuestas se han enfocado en modelos de predicción que incluyen estadísticas descriptivas, relaciones casuales, modelos de regresión y análisis de correlación[66], han llevado a que la epidemiología computacional haya surgido como alternativa para analizar gran cantidad de datos provenientes de múltiples fuentes de información. Dentro de los métodos emergentes se resaltan los modelos de red neuronal artificial y máquina de soporte vectorial como los más utilizados. Adicionalmente, la enfermedad arboviral más frecuentemente estudiada fue el dengue y la mayoría de los estudios se desarrollaron en Asia.

Aunque en diferentes partes del mundo se han desarrollado diversos modelos predictivos de enfermedades arbovirales utilizando la información de vigilancia, se ha reportado que las técnicas de modelado estadístico más utilizadas en los estudios de enfermedades arbovirales como el dengue son la regresión de Poisson, la NBM[58,67], ARIMA y GAM[68]. Los modelos ARIMA y GAM son modelos de referencia estándar para asociar factores ambientales con la enfermedad y una herramienta para el análisis de predicción de series de tiempo[43]. Sin embargo, aunque estas técnicas de modelado estadístico se utilizan ampliamente, tienen ciertas desventajas como el manejo de los valores perdidos, la sensibilidad de los valores atípicos y la multicolinealidad[63].

Otras alternativas que se han utilizado para predecir el comportamiento epidemiológico de las enfermedades arbovirales son los modelos espaciales bayesianos y las técnicas basadas en inteligencia artificial. Los primeros pueden reducir la varianza estimada de la variable respuesta, e incorporar componentes de la varianza en diferentes niveles[34], y las segundas tienen un enfoque no lineal, con menos supuestos sobre la distribución de los datos y las relaciones entre las variables[41]. Las anteriores características han permitido que estos modelos puedan considerar diversas fuentes de información y puedan tener una mejor predicción del evento. Sin embargo, las máquinas de vectores de soporte (SVM) podrían ser muy lentas para tareas a gran escala. También, a pesar del buen desempeño de los sistemas de pronóstico de redes neuronales, el modelo de red neuronal requiere una gran cantidad de datos y esfuerzo para ajustar sus hiperparámetros con el fin de evitar el sobreajuste. Por lo tanto, se necesita computación de alto rendimiento para el entrenamiento de los modelos.

En relación con las limitaciones para el uso de los métodos respecto a las fuentes de información, una de estas es la notificación “insuficiente”, esto es, los casos de dengue en los estudios están significativamente subregistrados, especialmente porque los casos leves se subnotifican y los asintomáticos no se diagnostican. Esto sesga la asociación con las variables predictoras estudiadas y ocasiona que los modelos subestimen la cantidad real de infecciones. Adicionalmente, en varios estudios, los recuentos oficiales de casos en algunas semanas eran inferiores a 50, mientras que en otras semanas epidémicas los casos superaban los 1000. Por lo tanto, los modelos no tuvieron la capacidad de predecir con exactitud el número de casos, sobreestimándolos en las semanas bajas y subestimándolos en las semanas altas. También se reconoce como limitación el periodo a predecir, puesto que la exactitud disminuye a media que la ventana de pronóstico es más larga.

En cuanto a las variables predictoras, una de las limitaciones es la omisión de algunas de estas por falta de información que obliga a ajustar estadísticamente los modelos, lo cual afecta su calidad y rendimiento. Dentro de estas variables se encuentran la densidad vectorial, la biología del vector, el movimiento humano, los servicios de salud, entre otras. Otra limitación para los estudios cuya base de información fueron las búsquedas en Internet sobre la enfermedad, es que las tasas de uso de Internet y de búsqueda de información cambiaron con el tiempo, y esto dificultó la estimación de los parámetros en los modelos. Adicionalmente, los patrones de búsqueda pueden reflejar la cobertura de los medios y el conocimiento de la situación epidemiológica por parte de la población puede no coincidir con la dinámica de la enfermedad que se está rastreando. Además, los diferentes países y ciudades tienen distintas maneras de reportar las noticias sobre estos eventos. De esta forma, los medios locales en regiones endémicas para enfermedades arbovirales pueden reaccionar de manera diferente a los brotes que en regiones no endémicas. Por lo tanto, la atención de los medios tiene el potencial de influir dramáticamente en las predicciones semanales.

Finalmente, la presente revisión se enfocó en identificar los métodos de aprendizaje automático usados para predecir el comportamiento epidemiológico de enfermedades arbovirales y documentar las variables predictoras utilizadas. Sin embargo, dentro de las limitaciones de esta revisión se reconoce que solo se utilizaron dos bases de datos de publicaciones científicas para la búsqueda de los artículos y que no se realizó búsqueda de literatura gris. De este modo, pudieron no haberse incluido algunos artículos relevantes para la revisión.

Conclusiones

En la última década se ha incrementado la publicación de trabajos de investigación que pretenden predecir el comportamiento epidemiológico de las enfermedades arbovirales utilizando diversos métodos de aprendizaje automático para apoyar a los tomadores de decisiones en salud pública.

Los diversos métodos permiten incorporar no solo las series de tiempo de los casos y las variables climatológicas, si no también otras fuentes de información incluida la movilidad de los humanos, y variables de datos abiertos como las consultas en la web y los comentarios en redes sociales[65]. El rendimiento predictivo de estos métodos puede variar dependiendo del área geográfica, la longitud de las series de tiempo, el periodo de la predicción, la enfermedad objeto de predicción y las variables predictoras utilizadas. De tal forma, la predicción puede ser precisa solo para el área en particular donde se realizó el estudio, no pudiéndose extender su aplicación a otras áreas o regiones. Por consiguiente, se requiere evaluar la precisión entre diversos modelos para seleccionar el que tenga mejor rendimiento de generalización del evento de interés en el área geográfica específica.

Agradecimientos

A la Universidad Industrial de Santander y a la Universidad de Santander por el financiamiento del proyecto **“**Modelos de analítica de datos para la vigilancia de enfermedades arbovirales en el departamento de Santander”, aprobado en el portafolio de la Vicerrectoría de Investigación y Extensión e identificado con el código 2694.

Conflicto de interés

Los autores declaramos que no tenemos conflictos de interés.

Referencias

1. World Health Organization. A Global Brief on Vector-Borne Diseases. Geneva: WHO; 2014. Disponible en: https://apps.who.int/iris/bitstream/handle/10665/111008/WHO_DCO_WHD_2014.1_eng.pdf
2. Warpeha KM, Munster V, Mullié C, Chen SH. Editorial: Emerging infectious and vector-borne diseases: A global challenge. Front. Public Health [Internet]. 2020; 8 (214): 1-2. doi: https://doi.org/10.3389%2Ffpubh.2020.00214
3. Wilder-Smith A, Gubler DJ, Weaver SC, Monath TP, Heymann DL, Scott TW. Epidemic arboviral diseases: priorities for research and public health. Lancet Infect Dis [Internet]. 2017; 17(3): e101-e106. doi: https://doi.org/10.1016/s1473-3099(16)30518-7
4. Padilla JC, Lizarazo FE, Murillo OL, Mendigaña FA, Pachón E, Vera MJ. Epidemiología de las principales enfermedades transmitidas por vectores en Colombia, 1990-2016. Biomédica (Bogotá) [Internet]. 2017; 37(Sup2): 27-40. doi: https://doi.org/10.7705/biomedica.v37i0.3769
5. Instituto Nacional de Salud de Colombia. Vectores de Dengue – Chikungunya , Estado Actual. Bogotá: INS; 2014. Disponible en: https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/IA/

INS/Zika-vector-22-mayo-2015-entomologia-vector.pdf
6. Mora-Salamanca AF, Porras-Ramírez A, De la Hoz Restrepo FP. Estimating the burden of arboviral diseases in Colombia between 2013 and 2016. Int J Infect Dis [Internet]. 2020; 97: 81-89. doi: https://doi.org/10.1016/j.ijid.2020.05.051
7. Hall HI, Correa A, Yoon PW, Braden CR, Centers for Desease Control and Prevention. Lexicon, definitions, and conceptual rramework for public health surveillance. MMWR Suppl. 2012; 61(3): 10-14.
8. Govindaraju V, Raghavan V, Rao CR. Handbook of statistics big data analytics. 2015. Vol 33.
9. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. Int J Inf Manage [Internet]. 2015; 35(2): 137-144. doi: https://doi.org/10.1016/j.ijinfomgt.2014.10.007
10. Martínez Sesmero JM. "Big data"; application and use for the health system. Farm Hosp [Internet]. 2015; 39(2): 69-70. doi: http://dx.doi.org/10.7399/fh.2015.39.2.8835
11. Mcafee A, Brynjolfsson E. Spotlight on big data big data: The management revolution. Harv Bus Rev. 2012; (October): 1-9. Disponible en: http://tarjomefa.com/wp-content/uploads/2017/04/6539-English-TarjomeFa-1.pdf
12. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffman TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021; 372(71). doi: https://doi.org/10.1136/bmj.n71
13. Goodman LB, Whittaker GR. Public health surveillance of infectious diseases: beyond point mutations. The Lancet Microbe. 2021; 2(2): e53-e54. doi: https://doi.org/10.1016/S2666-5247(21)00003-3
14. Bowman LR, Runge-Ranzinger S, McCall PJ. Assessing the Relationship between Vector Indices and Dengue Transmission: A Systematic Review of the Evidence. PLoS Negl Trop Dis. 2014; 8(5). doi: https://doi.org/10.1371/journal.pntd.0002848
15. Benedum CM, Seidahmed OME, Eltahir EAB, Markuzon N. Statistical modeling of the effect of rainfall flushing on dengue transmission in Singapore. PLoS Negl Trop Dis [Internet]. 2018; 12(12): 1-18. doi: https://doi.org/10.1371/journal.pntd.0006935
16. Carroll L, Au A, Detwiler LT, Fu T, Painter I, Abernethy N. Visualization and analytics tools for infectious disease epidemiology: A systematic review. J Biomed Inf [Internet]. 2014; 51: 287-298. doi: https://doi.org/10.1016/j.jbi.2014.04.006
17. Dwolatzky B, Trengove E, Struthers H, McIntyre JA, Martinson NA. Linking the global positioning system (GPS) to a personal digital assisstant (PDA) to support tuberculosis control in South Africa: A pilot study. Int J Health Geogr [Internet]. 2006; 5(34): 1-6. doi: https://doi.org/10.1186/1476-072x-5-34
18. Janies DA, Treseder T, Alexandrov B, Habib F, Chen JJ, Ferreira R,,et al. The Supramap project: Linking pathogen genomes with geography to fight emergent infectious diseases. Cladistics [Internet]. 2011; 27(1): 61-66. doi: https://doi.org/10.1111%2Fj.1096-0031.2010.00314.x
19. Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, Gangavarapu K, et al. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. Nature [Internet]. 2017; 546: 401-405. doi: https://doi.org/10.1038%2Fnature22400
20. Dellicour S, Rose R, Faria NR, Vieira LFP, Bourhy H, Gilbert M, et al. Using viral eene sequences to compare and explain the heterogeneous spatial dynamics of virus epidemics. Mol Biol Evol [Internet]. 2017; 34(10): 2563-2571. doi: https://doi.org/10.1093/molbev/msx176
21. Hansen TE, Hourcade JP, Segre A, Hlady C, Polgreen P, Wyman C. Interactive visualization of hospital contact network data on multi-touch displays. Proc 3rd Mex Work Hum Comput Interact. 2010; 1: 15-22.
22. Althouse BM, Ng YY, Cummings DAT. Prediction of dengue incidence using search query surveillance. PLoS Negl Trop Dis [Internet]. 2011; 5(8): 1-7. doi: https://doi.org/10.1371%2Fjournal.pntd.0001258
23. Teng Y, Bi D, Xie G, Jin Y, Huang Y, Lin B, et al. Dynamic forecasting of zika epidemics using google trends. PLoS One [Internet]. 2017; 12(1): 1-10. doi: https://doi.org/10.1371/journal.pone.0165085
24. Daughton AR, Paul MJ. Identifying protective health behaviors on twitter: Observational study of travel advisories and Zika virus. J Med Internet Res [Internet]. 2019; 21(5): 13090. doi: https://doi.org/10.2196%2F13090
25. McGough SF, Brownstein JS, Hawkins JB, Santillana M. Forecasting Zika incidence in the 2016 Latin America outbreak combining traditional disease surveillance with search, social media, and news report data. PLoS Negl Trop Dis [Internet]. 2017;11(1):1-15. doi: https://doi.org/10.1371/journal.pntd.0005295
26. Nsoesie EO, Flor L, Hawkins J, Maharana A, Skotnes T, Marinho F, et al. Social media as a sentinel for disease surveillance: What does sociodemographic status have to do with it? Plos Curr [Internet].

2016;7. doi: https://doi.org/10.1371%2Fcurrents.outbreaks.cc09a42586e16dc7dd62813b7ee5d6b6

27. Flahault A, Geissbuhler A, Guessous I, Guérin PJ, Bolon I, Salathé M, et al. Precision global health in the digital age. Swiss Med Wkly [Internet]. 2017;147. doi: https://doi.org/10.4414/smw.2017.14423
28. Chuang TW, Wimberly MC. Remote Sensing of Climatic Anomalies and West Nile Virus Incidence in the Northern Great Plains of the United States. PLoS One [Internet]. 2012; 7(10): 1-10. doi: https://doi.org/10.1371/journal.pone.0046882
29. Ruiz-Moreno D. Assessing Chikungunya risk in a metropolitan area of Argentina through satellite images and mathematical models. BMC Infect Dis [Internet]. 2016; 16(1): 1-12. doi: https://doi.org/10.1186/s12879-016-1348-y
30. Wu CH, Kao SC, Shih CH, Kan MH. Open data mining for Taiwan's dengue epidemic. Acta Trop [Internet]. 2018; 183: 1-7. doi: https://doi.org/10.1016/j.actatropica.2018.03.017
31. Guo P, Liu T, Zhang Q, Wang L, Xiao J, Zhang Q, et al. Developing a dengue forecast model using machine learning: A case study in China. PLoS Negl Trop Dis [Internet]. 2017; 11(10): 1-22. doi: https://doi.org/10.1371/journal.pntd.0005973
32. Baquero OS, Santana LMR, Chiaravalloti-Neto F. Dengue forecasting in São Paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models. PLoS One [Internet]. 2018; 13(4): 1-12. doi: https://doi.org/10.1371/journal.pone.0195065
33. Jain R, Sontisirikit S, Iamsirithaworn S, Prendinger H. Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data. BMC Infect Dis [Internet]. 2019; 19(1): 1-16. doi: https://doi.org/10.1186%2Fs12879-019-3874-x
34. Aswi A, Cramb SM, Moraga P, Mengersen K. Bayesian spatial and spatio-temporal approaches to modelling dengue fever: A systematic review. Epidemiol Infect [Internet]. 2019; 147: 33. doi: https://doi.org/10.1017/s0950268818002807
35. Akter R, Hu W, Gatton M, Bambrick H, Cheng J, Tong S. Climate variability, socio-ecological factors and dengue transmission in tropical Queensland, Australia: A Bayesian spatial analysis. Environ Res [Internet]. 2021; 195: 110285. doi: https://doi.org/10.1016/j.envres.2020.110285
36. Ho SH, Speldewinde P, Cook A. Predicting arboviral disease emergence using Bayesian networks: A case study of dengue virus in Western Australia. Epidemiol Infect [Internet]. 2017; 145(1): 54-66. doi: https://doi.org/10.1017/S0950268816002090
37. Martínez-Bello D, López-Quílez A, Prieto AT. Spatiotemporal modeling of relative risk of dengue disease in Colombia. Stoch Environ Res Risk Assess [Internet]. 2017; 32(6): 1587-1601. doi: https://doi.org/10.1007/s00477-017-1461-5
38. Deo R. Machine Learning in Medicine. Circulation [Internet]. 2015; 132(20): 1920–1930. doi: https://doi.org/10.1161/CIRCULATIONAHA.115.001593
39. Zhao N, Charland K, Carabali M, Nsoesie EO, Maheu-Giroux M, Rees E, et al. Machine learning and dengue forecasting : Comparing random forests and artificial neural networks for predicting dengue burden at national and sub-national scales in Colombia. PLoS Negl Trop Dis [Internet]. 2020; 14(9): 1-16. doi: https://doi.org/10.1371/journal.pntd.0008056
40. Liu K, Zhang M, Xi G, Deng A, Song T, Li Q, et al. Enhancing fine-grained intra-urban dengue forecasting by integrating spatial interactions of human movements between urban regions. PLoS Negl Trop Dis [Internet]. 2020; 14(12): 1-22. doi: https://doi.org/10.1371/journal.pntd.0008924
41. Sippy R, Farrell DF, Lichtenstein DA, Nightingale R, Harris MA, Toth J, et al. Severity index for suspected arbovirus (SISA): Machine learning for accurate prediction of hospitalization in subjects suspected of arboviral infection. PLoS Negl Trop Dis [Internet]. 2020; 14(2): 1-20. doi: https://doi.org/10.1371/journal.pntd.0007969
42. Jiang D, Hao M, Ding F, Fu J, Li M. Mapping the transmission risk of Zika virus using machine learning models. Acta Trop [Internet]. 2018; 185: 391-399. doi: https://doi.org/10.1016/j.actatropica.2018.06.021
43. Salim NAM, Wah YB, Reeves C, Smith M, Yaacob WFW, Mudin RN, et al. Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. Sci Rep [Internet]. 2021; 11(1): 1-9. doi: https://doi.org/10.1038/s41598-020-79193-2
44. Mussumeci E, Coelho FC. Large-scale multivariate forecasting models for Dengue - LSTM versus random forest regression. Spat Spatiotemporal Epidemiol [Internet]. 2020;35:100372. doi: https://doi.org/10.1016/j.sste.2020.100372
45. Xu J, Xu K, Li Z, Meng F, Tu T, Xu L, et al. Forecast of dengue cases in 20 chinese cities based on the deep learning method. Int J Environ Res Public Health [Internet]. 2020; 17(2): 453. doi: https://doi.org/10.3390/ijerph17020453
46. Bomfim R, Pei S, Shaman J, Yamana T, Makse HA, Andrade JS Jr, et al. Predicting dengue outbreaks at neighbourhood level using human mobility in urban

areas. Interface [Internet]. 2020; 17(171): 20200691. doi: http://dx.doi.org/10.1098/rsif.2020.0691

47. Akhtar M, Kraemer MUG, Gardner LM. A dynamic neural network model for predicting risk of Zika in real time. BMC Med [Internet]. 2019; 17(1). doi: http://dx.doi.org/10.1186/s12916-019-1389-3
48. Starczewski JT. Efficient triangular type-2 fuzzy logic systems. Int J Approx Reason [Internet]. 2009; 50(5): 799-811. doi: http://dx.doi.org/10.1016/j.ijar.2009.03.001
49. Torres C, Barguil S, Melgarejo M, Olarte A. Fuzzy model identification of dengue epidemic in Colombia based on multiresolution analysis. Artif Intell Med [Internet]. 2014; 60(1): 41-51. doi: http://dx.doi.org/10.1016/j.artmed.2013.11.008
50. Adak S, Jana S. A model to assess dengue using type 2 fuzzy inference system. Biomed Signal Process Control [Internet]. 2021; 63:102121. doi: http://dx.doi.org/10.1016/j.bspc.2020.102121
51. Wu CH, Kao SC. Knowledge discovery in open data for epidemic disease prediction. Heal Policy Technol [Internet]. 2021; 10(1): 126-134. doi: http://dx.doi.org/10.1016/j.hlpt.2021.01.001
52. Stolerman LM, Maia PD, Nathan Kutz J. Forecasting dengue fever in Brazil: An assessment of climate conditions. PLoS One [Internet]. 2019; 14(8). doi: http://dx.doi.org/10.1371/journal.pone.0220106
53. Chen Y, Ong JHY, Rajarethinam J, Yap G, Ng LC, Cook AR. Neighbourhoodlevel real-time forecasting of dengue cases in tropical urban Singapore. BMC Med [Internet]. 2018; 16(1): 1-13. doi: http://dx.doi.org/10.1186/s12916-018-1108-5
54. Shi Y, Liu X, Kok SY, Rajarethinam J, Liang S,Yap G, et al. Three-month real-time dengue forecast models: An early warning system for outbreak alerts and policy decision support in Singapore. Environ Health Perspect. 2016; 124(9): 1369-1375. doi: http://dx.doi.org/10.1289/ehp.1509981
55. Campbell KM, Haldeman K, Lehnig C, Munayco CV, Halsey ES, Laguna-Torres VA, et al. Weather regulates location, timing, and intensity of dengue virus transmission between humans and mosquitoes. PLoS Negl Trop Dis [Internet]. 2015; 9(7): 1-26. doi: https://doi.org/10.1371/journal.pntd.0003957
56. Ong J, Liu X, Rajarethinam J, Kok SY, Liang S, Tang CS, et al. Mapping dengue risk in Singapore using Random Forest. PLoS Negl Trop Dis [Internet]. 2018; 12(6): 1-12. doi: https://doi.org/10.1371/journal.pntd.0006587
57. Polwiang S. The time series seasonal patterns of dengue fever and associated weather variables in Bangkok (2003-2017). BMC Infect Dis [Internet]. 2020; 20(1): 1-10. doi: https://doi.org/10.1186/s12879-020-4902-6
58. Tuladhar R, Singh A, Banjara MR, Gautam I, Dhimal M, Varma A, et al. Effect of meteorological factors on the seasonal prevalence of dengue vectors in upland hilly and lowland Terai regions of Nepal. Parasites and Vectors [Internet]. 2019; 12(1):1-15. doi: https://doi.org/10.1186/s13071-019-3304-3
59. Yuan HY, Liang J, Lin PS, Sucipto K, Tsegaye MM, Wen TH, et al. The effects of seasonal climate variability on dengue annual incidence in Hong Kong: A modelling study. Sci Rep [Internet]. 2020; 10(1): 1-10. doi: https://doi.org/10.1038/s41598-020-60309-7
60. Nejad YF, Varathan KD. Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction. BMC Med Inform Decis Mak [Internet]. 2021; 21(1): 1-12. doi: https://doi.org/10.1186/s12911-021-01493-y
61. McGough SF, Clemente L, Kutz JN, Santillana M. A dynamic, ensemble learning approach to forecast dengue fever epidemic years in Brazil using weather and population susceptibility cycles. J R Soc Interface [Internet]. 2021; 18(179): 20201006. doi: https://doi.org/10.1098/rsif.2020.1006
62. Benedum CM, Shea KM, Jenkins HE, Kim LY, Markuzon N. Weekly dengue forecasts in Iquitos, Peru; San Juan, Puerto Rico; and Singapore. PLoS Negl Trop Dis [Internet]. 2020; 14(10): 1-26. doi: https://doi.org/10.1371/journal.pntd.0008710
63. Carvajal TM, Viacrusis KM, Hernandez LFT, Ho HT, Amalin DM, Watanabe K. Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. BMC Infect Dis [Internet]. 2018; 18(1): 1-15. doi: https://doi.org/10.1186/s12879-018-3066-0
64. Liu K, Yin L, Zhang M, Kang M, Deng AP, Li QL, et al. Facilitating fine-grained intra-urban dengue forecasting by integrating urban environments measured from street-view images. Infect Dis Poverty [Internet]. 2021;10(40):1-16. doi: https://doi.org/10.1186/s40249-021-00824-5
65. Liu X, Rajarethinam J, Shi Y, Liang S, Yap G, Ng LC. Development of predictive dengue risk map using Random Forest. Int J Infect Dis [Internet]. 2016; 45: 346. doi: https://doi.org/10.1016/j.ijid.2016.02.746
66. Hsu JC, Hsieh CL, Lu CY. Trend and geographic analysis of the prevalence of dengue in Taiwan, 2010–2015. Int J Infect Dis [Internet]. 2017; 54: 43-49. doi: https://doi.org/10.1016/j.ijid.2016.11.008
67. Ho TS, Weng TC, Wang JD, Han HC, Cheng HC,

Yang CC, et al. Comparing machine learning with case- control models to identify confirmed dengue cases. PLoS Negl Trop Dis [Internet]. 2020; 14(11): 1-21. doi: https://doi.org/10.1371/journal.pntd.0008843

68. Xiao J, Liu T, Lin H, Zhu G, Zeng W, Li X, et al. Weather variables and the El Niño Southern Oscillation may drive the epidemics of dengue in Guangdong Province, China. Sci Total Environ [Internet]. 2018; 624: 926-934. doi: https://doi.org/10.1016/j.scitotenv.2017.12.200